PERGAMON Mathematical and Computer Modelling 32 (2000) 273–288

MATHEMATICAL AND COMPUTER MODELLING

www.elsevier.nl/locate/mcm

Classical and Bayesian Approaches to Compartment Models Based on *in vivo* Cadmium Data

S. L. SHENG AND N. BALAKRISHNAN
Department of Mathematics and Statistics, McMaster University
Hamilton, Ontario, Canada L8S 4K1

D. R. CHETTLE
Department of Engineering Physics, McMaster University
Hamilton, Ontario, Canada L8S 4L7

Abstract—Four-compartment models in the form of a system of linear differential equations, associated with *in vivo* cadmium data, have been constructed to describe a particular biological phenomenon of cadmium metabolism in the human body. Since it is quite technically expensive to make *in vivo* measurements and also difficult to collect data over a period of a decade, the data analyzed in this paper are quite precious, and in fact, one of very few such existing data sets in the area of cadmium research. Cadmium researchers are very much interested in drawing as much information as possible from these data. Methods are developed here for deriving the expectation functions and for using them to analyze this special data set. The parameter estimation for compartment models discussed next uses classical as well as Bayesian approaches. This is the first time that a whole system of the human body is analyzed simultaneously, instead of discussing each compartment separately with a large number of assumptions from different sources on Kjellström's model, without any additional assumption on the derived compartment models. The results, obtained from these statistical approaches on simple as well as more direct mathematical models, are not only interpretable very reasonably in terms of biological phenomena, but also reveal interesting patterns and show a great consistency with previous studies.

Keywords—Cadmium metabolism, Compartment model, Bayesian estimation, Parameter estimation.

1. INTRODUCTION

Cadmium has been used in a wide range of industrial applications even though its toxic effects have been acknowledged for decades [1]. Toxic effects of cadmium occur if some significant biological mechanism is stimulated beyond a certain level. Of special interest among chronic effects of cadmium is kidney damage (being the most typical feature of chronic cadmium intoxication), which is characterized by an excess excretion of low molecular weight proteins such as β_2-microglobulin [2].

Cadmium researchers are therefore quite interested in biological phenomenon of cadmium metabolism, and in particular with the absorption, distribution, accumulation, and excretion of cadmium with respect to time in the human body. The absorption and subsequent distribution of cadmium depend on the particular exposure route and the form and intensity to which

0895-7177/00/$ - see front matter Typeset by $\mathcal{A}\mathcal{M}\mathcal{S}$-TEX
PII: S0895-7177(00)00134-5

workers are exposed. Cadmium is usually inhaled or ingested and, initially, a considerable amount will be found in lungs or intestinal wall from where it is transported, by blood plasma, to other tissues. Cadmium in blood plasma is bound to low molecular weight as well as high molecular weight proteins, and it is the transport and metabolism of these that determines, to a large extent, the distribution of cadmium through the body. The main cadmium accumulation organs are the liver and kidneys, which account for approximately 16% and 53% of the total body burden, respectively [2,3]. Accumulation in these sites results from the binding of cadmium to proteins—particularly, albumin and metallothionein for liver, and metallothionein for the kidney. An eight-compartment kinetic model, showing a flow scheme of cadmium metabolism, has been established. This model, with 21 distribution coefficients, contains compartments of liver, kidney, feces, urine, other tissues, blood 1, blood 2, and blood 3 cadmium. Blood 1 contributes to accumulation of cadmium in liver and other tissues, while blood 3 contributes to accumulation of cadmium in kidneys. Blood 2, between blood 1 and blood 3, contains the accumulation of cadmium bound to cells and molecules. Background information for Kjellström's model had been obtained partly from animal experiments and partly from observations on the human beings in industrial and general environmental exposure situations.

In addition to a number of assumptions in Kjellström's model, all the unknown parameters are fixed according to prior experience and knowledge based on different sources and different animal and autopsy data sets. During the analysis, the numerical solution of only one compartment of the model can be derived after the unknown parameters are fixed. Thus, the relationship between each compartment and time can be established with prior fixed parameters. In contrast to Kjellström's model, the four-compartment model discussed in this paper is formulated with systems of linear differential equations in a simple and more direct manner. The parameter estimation for the four-compartment model is discussed using classical and Bayesian approaches. This is the first time that a whole system of the human body is analyzed simultaneously without any additional assumption on the derived compartment models; usually, discussions are made on each compartment separately, with a number of assumptions concerning the unknown parameters. Furthermore, all the parameters are estimated from observed human data in the four-compartment model.

Kjellström's model provides a good biological basis and reference for the starting values of the unknown parameters while dealing with the four-compartment model, even though the parameter estimation based on the four-compartment model is achieved in an inverse way mathematically. The results obtained here are not only interpretable very reasonably in terms of biological phenomena, but also reveal interesting patterns and show a great consistency with previous studies, although the fact that data here are sparse and noisy brings about the special difficulty encountered in this study.

2. PROBLEM FORMULATION

2.1. Subjects

The subjects of this paper are 14 male workers involved in brazing using cadmium-silver alloys. The duration of cadmium exposure ranged from 2 to 34 years. The level of exposure for each subject varied throughout the exposure time according to the work undertaken. The exposure for all but one subject (number 3) was at or above the current U.K. occupational limit for cadmium in air, i.e., 0.05 mg/m^3, for at least part of the time.

Longitudinal measurements including biochemical and *in vivo* measurements were made for these 14 male workers. Biochemical measurements consisting of urinary cadmium, blood cadmium, and urinary β_2-microglobulin had been made at approximately six monthly to yearly intervals from 1983 to 1990. Blood and urinary cadmium were measured by using electrothermal atomic spectrophotometry, while the urinary β_2-microglobulin was measured in freshly voided samples using a radioimmunoassay procedure.

In addition to the biochemical measurements, all subjects had *in vivo* measurements of liver and kidney cadmium made in December 1983, and eight subjects (numbers 1, 2, 3, 4, 5, 10, 11, 13) had additional *in vivo* measurements made in March/April 1990. These organ levels were measured by prompt γ-ray neutron activation analysis, using different measurement systems on each occasion. The differences arising solely from the different measurement systems are thought to be negligible.

In vivo measurement of the levels of liver and kidney cadmium along with urine and blood cadmium, as well as low molecular weight proteins β_2-microglobulin provide estimates in order to describe the phenomenon of cadmium metabolism, based on the compartment model [4]. Since it is quite expensive to make *in vivo* measurements and also difficult to collect data over a period of decades, the data analyzed in this paper are quite precious, and in fact one of very few such existing data sets in the area of cadmium research. Cadmium researchers are interested in drawing as much information as possible from these data.

In this paper, the abbreviations Cd_u, Cd_b, Cd_l, Cd_k, and u_β_2 are used in some formulae, which denote measurements of urine cadmium, blood cadmium, liver cadmium, kidney cadmium, and urinary β_2-microglobulin, respectively. The duration of cadmium exposure, the first measurement and last measurement of u_β_2 as an indicator of kidney status, the history of smoking and age for each subject are listed in Table 1.

Table 1. Duration of exposure, age, and indicator of kidney status.

Subject Number	Cadmium Exposure (years)	Year of Birth	*β_2-Microglobulin (mg/g creat.)	Smoking Habit
1	2	1964	257/NA	not smoking
2	31	1937	90/57	not smoking
3	13	1945	58/135	not smoking
4	33	1930	86/238	smoking (1948–)
5	20	1928	688/88	not smoking
6	9	1923	344/NA	NA
7	3	1961	19/NA	NA
8	2	1925	59/NA	NA
9	24	1928	1667/NA	NA
10	34	1921	3077/84600	smoking (1939–1958)
11	16	1920	158000/154100	smoking (1938–1962)
12	22	1927	13400/155800	NA
13	22	1925	1067/NA	smoking (1941–1975)
14	24	1924	37000/102500	NA

In the 'β_2-microglobulin' column, the left side of backslash is for the first measurement, and the right side is for the last one.

'NA' means that the information is not available.

2.2. Compartment Model

Derivation of the solution of the four-compartment model, a linear system of ordinary differential equations, is discussed in this section, in which the four compartments are cadmium in blood, cadmium in urine, cadmium in kidney, and cadmium in liver. It is assumed that the rates of

flow of cadmium follow the first-order kinetics, in which the mass balance equations are required for the description of dynamics of the exchange of cadmium among compartments. The transfer coefficient is assumed to be constant with respect to time. The model can then be described analytically in the form of a linear system of ordinary differential equations as follows:

$$\frac{dH}{dt} = AH + B, \qquad t \geq 0,$$
$$H(0) = H_0,$$

where

$$H = \begin{bmatrix} \eta_b \\ \eta_l \\ \eta_k \\ \eta_u \end{bmatrix}, \qquad A = \begin{bmatrix} -(C_{bl} + C_{bk}) & C_{lb} & C_{kb} & 0 \\ C_{bl} & -C_{lb} & 0 & 0 \\ C_{bk} & 0 & -(C_{kb} + C_{ku}) & 0 \\ 0 & 0 & C_{ku} & 0 \end{bmatrix}, \qquad B = \begin{bmatrix} R \\ 0 \\ 0 \\ 0 \end{bmatrix},$$

with the parameters C_{bl}, C_{lb}, C_{bk}, C_{kb}, and C_{ku} in the matrix A denoting the transfer rates of cadmium from blood to liver, from liver to blood, from blood to kidney, from kidney to blood, and from kidney to urine, respectively; R in the vector B stands for the monthly intake of cadmium absorbed into blood from environment; the elements of H, viz. η_b, η_l, η_k, and η_u, denote the expected responses of cadmium in blood, liver, kidney, and urine, respectively; t is time in monthly unit; and each element of H_0 denotes the initial value of corresponding element of H at time 0.

The compartment system can also be represented graphically, in which the boxes represent compartments, and the arrows labeled with transfer coefficients represent transfers of material into or out of compartments. The compartment diagram for the model described above is shown in Figure 1.

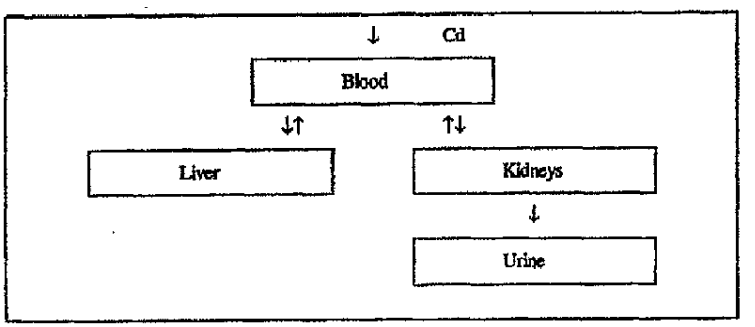

Figure 1. Diagram for the four-compartment model (Cd_b, Cd_l, Cd_k, and Cd_u).

Since the measurements on the amount of cadmium in urine per month are available and urine cadmium has little feedback to kidney, the dimension can be reduced to take the first three ordinary differential equations into consideration at the initial stage of deriving equations for the expected responses. For simplicity in computation, we shall use

$$\frac{dH_1}{dt} = A_1 H_1 + B_1, \qquad t \geq 0,$$
$$H_1(0) = H_{1_0},$$
$$\frac{d\eta_u}{dt} = C_{ku}\eta_k,$$

where

$$H_1 = \begin{bmatrix} \eta_b \\ \eta_l \\ \eta_k \end{bmatrix}, \qquad A_1 = \begin{bmatrix} -(C_{bl} + C_{bk}) & C_{lg} & C_{kb} \\ C_{bl} & -C_{lb} & 0 \\ C_{bk} & 0 & -(C_{kb} + C_{ku}) \end{bmatrix}, \qquad B_1 = \begin{bmatrix} R \\ 0 \\ 0 \end{bmatrix}$$

The common approach for constructing the expectation functions is to derive the weighted sum-of-exponential model in which the expected responses can be expressed as combinations of weighted sum of exponentials and the constants due to the special form of the ordinary differential equations (without solving them). From this approach, one cannot directly estimate the parameters in the generating system, and the number of exponential parameters may exceed the number of system parameters. It is not practical to use this approach for setting the functions of expected responses on the cadmium data we have here, since the number of measurements for each subject is at most 12, which is not large enough, and the data structure is irregular with some missing values.

We, therefore, adopt here a combination of analytical approach and exponential matrix approach. Under this approach, the expectation functions for the blood, liver, and kidney cadmium can be expressed by elements of A_1, the elements of B_1, the eigenvalues of A_1, and the elements of the corresponding eigenvectors.

In order to find all the eigenvalues of A_1 and the corresponding eigenvectors, consider the characteristic polynomial of A_1 given by

$$p(\lambda) = \det(A_1 - \lambda I_3),$$

where $\det(A_1 - \lambda I_3)$ denotes the determinant of matrix $A_1 - \lambda I_3$, I_3 denotes a 3×3 identity matrix, and λ is a scalar. This characteristic polynomial can be simplified as

$$\lambda^3 + m_1\lambda^2 + m_2\lambda + m_3 = 0,$$

where

$$\begin{aligned} m_1 &= C_{bl} + C_{lb} + C_{bk} + C_{kb} + C_{ku}, \\ m_2 &= C_{bl}C_{kb} + C_{bl}C_{ku} + C_{bk}C_{kb} + C_{bk}C_{ku} + C_{lb}C_{ku} + C_{lb}C_{bk}, \\ m_3 &= C_{lb}C_{bk}C_{ku}. \end{aligned}$$

To solve the equation easily, we make the change $\lambda = x - m_1/3$. It then turns out to be

$$x^3 + f_1 x + f_2 = 0,$$

where

$$\begin{aligned} f_1 &= -\frac{m_1^2}{3} + m_2, \\ f_2 &= \frac{2}{27} m_1^2 - \frac{m_1 m_2}{3} + m_3. \end{aligned}$$

By applying Cardan's formula [5], the discriminant Δ for the equation is given by

$$\Delta = \frac{1}{9} f_1^3 + \frac{1}{4} f_2^2.$$

There are four types of eigenvalues of A_1, depending on what the above discriminant Δ is.

Type 1

If Δ is greater than zero, only one eigenvalue is real and the other two are complex conjugate values. The expressions of eigenvalues in this case are given by

$$\begin{aligned} \lambda_1 &= \alpha + \beta - \frac{m_1}{3}, \\ \lambda_2 &= -\frac{1}{2}(\alpha + \beta) + \frac{\sqrt{3}}{2} i(\alpha - \beta) - \frac{m_1}{3}, \\ \lambda_3 &= -\frac{1}{2}(\alpha + \beta) - \frac{\sqrt{3}}{2} i(\alpha - \beta) - \frac{m_1}{3}, \end{aligned}$$

and

$$\alpha = \sqrt[3]{-\frac{f_2}{2} + \sqrt{\Delta}}; \qquad \beta = \sqrt[3]{-\frac{f_2}{2} - \sqrt{\Delta}},$$

where $i = \sqrt{-1}$ is the imaginary number.

Type 2

If Δ is equal to zero, one of the possibilities is that $f_1 = f_2 = 0$. All the eigenvalues are then the same and real in this case, and are given by

$$\lambda_1 = \lambda_2 = \lambda_3 = -\frac{m_1}{3}.$$

Type 3

If Δ is equal to zero, another possibility is that $f_1^2/9 = -f_2^2/4$. In this case, two eigenvalues are the same and all of them are real. They are given by

$$\lambda_1 = \alpha + \beta - \frac{m_1}{3}$$

and

$$\lambda_2 = \lambda_3 = -\frac{(\alpha+\beta)}{2} - \frac{m_1}{3}.$$

Type 4

If Δ is less than zero, the eigenvalues are all distinct. These eigenvalues and the corresponding eigenvectors are given by

$$\lambda_j = 2\sqrt{-\frac{f_1}{2}}\cos\left[\frac{1}{3}\cos\left(-\frac{f_2}{2}\sqrt{-\frac{27}{f_1^3}}\right) + 2j\pi\right] - \frac{m_1}{3}, \qquad j = 1,2,3,$$

and

$$U = \begin{bmatrix} 1 & 1 & 1 \\ \dfrac{n_1}{\sqrt{1+n_1^2+p_1^2}} & \dfrac{n_2}{\sqrt{1+n_2^2+p_2^2}} & \dfrac{n_3}{\sqrt{1+n_3^2+p_3^2}} \\ \dfrac{p_1}{\sqrt{1+n_1^2+p_1^2}} & \dfrac{p_2}{\sqrt{1+n_2^2+p_2^2}} & \dfrac{p_3}{\sqrt{1+n_3^2+p_3^2}} \end{bmatrix},$$

where

$$n_i = \frac{C_{bl}}{\lambda_i + C_{lb}} \quad \text{and} \quad p_i = \frac{C_{bk}}{\lambda_i + C_{kb} + C_{ku}}, \qquad i = 1,2,3.$$

Biologically, it will not be meaningful to have Type I since it will not be meaningful for the eigenvalues, associated with the half-life time of cadmium in blood, kidney, and liver, to be complex values. We can rule out Type 2 as well, since it is not realistic to have all three eigenvalues be the same, as the half-life time of blood cadmium cannot be the same as the half-life time of either liver cadmium or kidney cadmium. There is a diminutive possibility of having Type 3 due to the uncertainty in the parameter estimation and the possible similarity between the half-life time of kidney cadmium and that of liver cadmium. Therefore, it is most likely to be of Type 4 for all subjects in that three eigenvalues derived from matrix A_1 are all distinct.

The solutions of linear systems of ordinary differential equations with initial values exist in the defined region and can be formed in terms of matrix exponential defined as $\exp(A_1 t) = \sum_{k=0}^{\infty}(A_1^k t^k/k!)$; see [6]. The matrix exponential can be written as

$$\exp(A_1 t) = U\exp(Jt)U^{-1},$$

where λ_i $(i = 1,\ldots,n)$ are the eigenvalues of A_1, U is nonsingular, the columns of U are the corresponding eigenvectors of A_1, and $\exp(Jt)$ is $\text{diag}\{\exp(\lambda_1 t),\ldots,\exp(\lambda_n t)\}$.

It is assumed that the three eigenvalues λ_i $(i = 1, 2, 3)$ of A_1 are dissimilar in each subject. The solutions of the compartment model as expectation functions can then be derived as

$$\begin{aligned} H_1(t) &= \exp(A_1 t)\, H_{1_0} + \exp(A_1 t) \int_0^t \exp^{-1}(A_1 \eta)\, B_1\, d\eta \\ &= U \exp(Jt) U^{-1} H_{1_0} + U \exp(Jt) U^{-1} \left(I - U \exp(-Jt) U^{-1}\right) A_1^{-1} B_1 \\ &= U \exp(Jt) U^{-1} \left(H_{1_0} + A_1^{-1} B_1\right) - A_1^{-1} B_1, \end{aligned}$$

where

$$\exp(Jt) = \begin{bmatrix} e^{\lambda_1 t'} & 0 & \\ 0 & e^{\lambda_2 t} & 0 \\ 0 & 0 & e^{\lambda_3 t} \end{bmatrix} \quad \text{and} \quad t' = \begin{cases} t, & t \geq 0, \\ 0, & t < 0. \end{cases}$$

The columns of U are the corresponding eigenvectors of A_1. The eigenvalues of A_1, λ_i $(i = 1, 2, 3)$, and the corresponding eigenvectors in the expectation functions which have been presented all correspond to Type 4.

If the situation in Type 3 takes place, the matrix $\text{diag}\{\exp(\lambda_1 t), \ldots, \exp(\lambda_n t)\}$ should be replaced by a general form, viz. the Jordan canonical form. The corresponding eigenvectors should be the generalized eigenvectors. This situation, however, will not be discussed in this paper.

Comparing the expectation functions to the ordinary differential equations, it can be noticed that the expectation functions have been modified when the time is less than 0. There are two possible reasons for the modification of the model.

1. Since the cadmium exposure to the workers ceased at time t_0 and the first measurement was made when the cadmium exposure already started, the level of cadmium in blood cannot be a dramatic change from the time of first measurement to time 0, and it would decrease right after the exposure stopped until it reaches another steady state.
2. There is no direct relationship between the intake of cadmium and liver cadmium, or kidney cadmium. The intake of cadmium has direct contact with blood cadmium through respiratory system and blood circulation system.

3. PARAMETER ESTIMATION

3.1. Two Methods of Estimation

Multivariate analysis has been used to estimate the unknown parameters with four responses (cadmium in blood, liver, kidney, and urine).

In general, the data for multiresponse parameter estimation consists of M observed responses, $y_{n1}, y_{n2}, \ldots, y_{nM}$, taken at each of N experimental conditions, x_n $(n = 1, 2, \ldots, N)$. These data are modeled as

$$y_{nm} = \eta_m(x_n, \theta) + \varepsilon_{nm}, \qquad n = 1, \ldots, N, \quad m = 1, \ldots, M,$$

where θ is a $P \times 1$ vector of unknown parameters; all observations y_{nm} are gathered into an $N \times M$ matrix Y; η_m is the model function of known form for the m^{th} response depending on some or all of the experimental setting x_n and on some or all of the parameters θ, and all $\eta_m(x_n, \theta)$ are collected into $N \times M$ expected response matrix $H(\theta)$; ε_{nm} is the disturbance term, which is assumed to be normally distributed and all the residuals are collected into an $N \times M$ matrix $Z(\theta) = Y(\theta) - H(\theta)$.

In the cadmium data, M is always 4, the number of unknown parameters P is 6, N varies from 9 to 12, and θ is the vector $[C_{bl}, C_{lb}, C_{bk}, C_{kb}, C_{ku}, R]'$.

Procedures for the determination of parameters in this multivariate system are based on two criteria, one based on the generalized nonlinear least-squares minimization, and the other based on the Bayesian determinant minimization.

The objective function of generalized nonlinear least-squares can be written as

$$h(\theta) = \sum_{m=1}^{M} \sum_{n=1}^{N} [y_{nm} - \eta_m (x_n, \theta)]^2 . \tag{3.1}$$

The following assumptions are now made:

1. all variances are known;
2. all variances are assumed equal;
3. there is no correlation between y_{ui} and y_{uj}, $i \neq j$ and $i, j = 1, \ldots, M$ and $u = 1, \ldots, N$.

If Assumption 2 alone is violated, appropriate weights could be added in the object. If, in addition to Assumption 2, Assumption 3 is also violated, besides adding weights, the cross terms (accounting for correlations) should also be included into the objective function.

In the Bayesian framework, the likelihood function (based on observations Y) for θ and Σ can be expressed as

$$\begin{aligned} l(\theta, \Sigma \mid Y) &\propto p(E \mid \Sigma, \theta) \\ &\propto |\Sigma|^{-N/2} \exp\left[-\frac{1}{2} \operatorname{tr} \Sigma^{-1} S(\theta)\right], \end{aligned}$$

where

$$\begin{aligned} S_{ij}(\theta_i, \theta_j) &= \sum_{u=1}^{n} \varepsilon_{ui} \varepsilon_{uj} \\ &= \sum_{u=1}^{n} [y_{ui} - \eta_i (\xi_{ui}, \theta_i)] [y_{uj} - \eta_j (\xi_{uj}, \theta_j)] \\ &= \varepsilon_i \varepsilon_j, \qquad i, j = 1, \ldots, m. \end{aligned}$$

Box and Tiao [7] used the Bayesian argument setting the prior distribution to be

$$p(\theta, \Sigma) = p(\theta) p(\Sigma) \propto |\Sigma|^{-(M+1)/2}$$

based on the assumption of independence between θ and Σ. The joint posterior distribution of (θ, Σ) is then

$$\begin{aligned} p(\theta, \Sigma \mid Y) &\propto p(\theta, \Sigma) l(\theta, \Sigma) \\ &\propto |\Sigma|^{-(N+M+1)/2} \exp\left[-\frac{1}{2} \operatorname{tr} \Sigma^{-1} S(\theta)\right], \qquad -\infty < \theta < \infty, \quad \Sigma > 0. \end{aligned}$$

It then follows that

$$p\left(\theta, \Sigma^{-1} \mid Y\right) \propto |\Sigma|^{(N-M-1)/2} \exp\left[-\frac{1}{2} \operatorname{tr} \Sigma^{-1} S(\theta)\right], \qquad -\infty < \theta < \infty, \quad \Sigma > 0.$$

Since Wishart distribution $W_m(B^{-1}, q)$ has its probability density function

$$W_m\left(B^{-1}, q\right) = p(Z) = k|Z|^{1/2(q-1)} \exp\left(-\frac{1}{2} \operatorname{tr} ZB\right), \qquad Z > 0,$$

where Z is distributed as Wishart with q degrees of freedom and parameter matrix B^{-1}, it can be shown that

$$\int_{Z>0} |Z|^{1/2(q-1)} \exp\left(-\frac{1}{2} \operatorname{tr} ZB\right) dZ = |B|^{-1/2(q+m-1)} 2^{1/2m(q+m-1)} \Gamma_m \left(\frac{q+m-1}{2}\right),$$

where

$$\Gamma_p(b) = \left[\Gamma\left(\frac{1}{2}\right)\right]^{1/2p(p-1)} \prod_{\alpha=1}^{p} \Gamma\left(b + \frac{\alpha - 1}{2}\right), \qquad b > \frac{1-\alpha}{2}.$$

Thus, the posterior density can be integrated over Σ to give

$$P(\theta \mid \Sigma) \propto \left|\Sigma^{\top}\Sigma\right|^{-N/2}.$$

As a result, the parameter estimates are to be chosen so as to minimize

$$d(\theta) = \left|\Sigma^{\top}\Sigma\right|. \tag{3.2}$$

The Bayesian determinant criterion $d(\theta)$ involves constraints on the number of observations (N), the number of responses (M), and the number of parameters (P). We need to have $N > P$ and $N \leq M$ in order to carry out the necessary computations; see [8].

3.2. Results

3.2.1. The first-stage analysis

The first-stage involves a two-compartment model which forms a preliminary work for the estimation of parameters of a four-compartment model to be considered later.

Since the monthly amount of urine cadmium is linearly related to kidney cadmium, the rate of change of expected kidney cadmium $\eta_k(t)$ at time t can be represented by

$$\frac{d\eta_k(t)}{dt} = -\lambda_k \eta_k(t) + R_k.$$

If kidney cadmium at time $t = 0$ is $\eta_k(0)$ and $\lambda_k \eta_k$ is the urinary output rate, the joint model for the urine and kidney cadmium is given by

$$\eta_k(t) = \frac{R_k}{\lambda_k}\left(1 - \exp\left(-\lambda_k t\right)\right) + \eta_k(0)\exp\left(-\lambda_k t\right),$$
$$\frac{d\eta_u(t)}{dt} = R_k\left(1 - \exp\left(-\lambda_k t\right)\right) + \lambda_k \eta_k(0)\exp\left(-\lambda_k t\right),$$

where η_k and η_u denote the expected responses of kidney and urine cadmium, respectively, λ_k is the rate from kidney cadmium to urine cadmium, and R_k is the monthly intake of cadmium from the environment into kidney. The diagram for the two-compartment model of kidney cadmium and urine cadmium is presented in Figure 2.

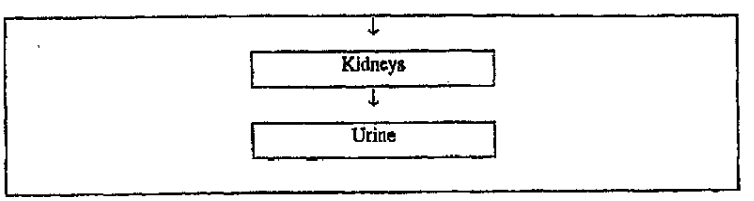

Figure 2. Diagram for the two-compartment model (Cd_u and Cd_k).

In the least-squares approach based on (3.1), $M = 2$ and θ is a 2×1 vector of unknown parameters, which is $[\lambda_k, R_k]^{\top}$. During the parameter estimation process, measurement of kidney cadmium at time 0, which is 12/1983, is added as an unknown parameter. The unit of monthly urine cadmium is converted by multiplying by 53 on the assumption that daily output of urine is 1.75 g. The results from the nonlinear least-squares estimation show that the range of estimates

of λ_k is from 0.0047 to 0.0537, which reveals that it is essential to include the input rate into the model. Furthermore, the results indicate that the standard deviations for most subjects are quite large, even though the estimates of kidney cadmium at time 0 and prediction of kidney cadmium at time 76 are rather close to the measured values. Since the observed values of kidney cadmium are quite dominant in the data so that the second assumption of the least-squares approach is violated, weights were added into the objective function.

In order to check the sensitivity of the weights used in the two-compartment model, several different ways were used to add weights into the model, based on prior experience with regard to the measurement error.

After weights were added, the residual sum of squares got dramatically reduced. It was interesting to observe that the estimate of λ_k was not really sensitive to the different methods of weights used. In Subject 7, the estimate of λ_k was the smallest in the model with every weight. The largest estimate of λ_k was located either in Subject 11 or 12 among the different weights. Table 2 shows one of the results from the weighted least-squares approach.

Table 2. Results of the nonlinear least-squares for the two-compartment model (Cd_k and Cd_u) based on weight.

Subject	$\hat{\lambda}_K$	$\hat{R}_K$	$\widehat{\text{Cd\_k}(0)}$	Cd_k(0) (S.)	$\widehat{\text{Cd\_k}(76)}$	Cd_k(76) (B.)	SSQ
1	0.007	40.09	2303.9	2500	5069.61	4000	6.933
2	0.0167	262.34	20965.8	21000	23079	24000	3.761
3	0.015	254.41	13795.7	12000	20360.6	24000	5.454
4	0.0069	193.22	15846.9	14000	30187.6	34000	8.449
5	0.0124	141.13	3786.7	2500	9897.49	12000	10.743
6	0.014	345.5	21041.6	21000	30684.4		5.506
7	0.0048	94.26	2472.3	2500	9435.7		4.679
8	0.0116	204.19	15012.9	15000	22747.3		3.230
9	0.0220	426.65	15253.9	15000	21481.3		2.206
10	0.0130	286.69	68846.2	78000	65108	24000	119.27
11	0.0271	1028.3	31154.9	31000	41051.3	42000	3.339
12	0.0495	675.8	10966.0	10000	13844.9		3.417
13	0.0150	681.9	24928.9	16000	46866.5	66000	53.046
14	0.0236	630.5	23958.0	24000	30242.9		2.258

S. refers to single.

In the Bayesian approach, some variables of kidney cadmium were simulated to satisfy the Bayesian constraint such that $N > P$ and $N \geq M$. It is assumed that the measurement errors of kidney cadmium are normally distributed. The generated variable of kidney cadmium at time t is provided by

$$\text{Cd\_k}(t) = n\sigma + \mu,$$

where n was generated to be normally distributed with mean zero and standard deviation one, the initial time t_0 is set at 0 and the unit of time is month, μ is the prediction of kidney cadmium at time t, derived from the measurements of kidney cadmium at time $t = 0$ and $t = 76$, and σ is the standard deviation of the measurement error in kidney cadmium, which is assumed to be

6000 μg. Therefore, the simulation is based on the measurements of kidney cadmium at time 0 and 76, standard normal variables n, the standard deviation of measurement error in kidney cadmium σ, and the prediction of kidney cadmium μ.

Table 3 shows the results from this Bayesian approach, in which the estimated λ_k for each subject falls in a reasonable range that is not too different from the results of the nonlinear least-squares. The variation of the estimate of R_k is quite large under both approaches. One of the reasons for this may be that the input rate is not directly related to urine cadmium but is connected directly to blood cadmium, and yet blood cadmium is not included in the two-compartment model.

Table 3. Results of the two-compartment model from the Bayesian approach with the combination of simulated and real data.

Subject	$\hat{\lambda}_K$	$\hat{R}_K$	$\widehat{Cd\_k(0)}$	Cd_k(0) (S.)	$\widehat{Cd\_k(76)}$	Cd_k(76) (B.)	O.F.V.
1	0.006	54.50	2759.63	2500	6824.4	4000	20.49
2	0.018	276.50	19169	21000	21211	24000	38.82
3	0.019	235.82	11520	12000	14920	24000	65.22
4	0.008	170.22	14754	14000	25759	34000	77.76
5	0.015	105.64	36780	2500	28316	12000	280.93
10	0.012	516.62	71838	78000	83474	24000	2009.22
11	0.027	990.71	31958	31000	40191	42000	21.4776
13	0.027	275.53	19518	16000	13909	66000	525.02

In order to obtain more precise estimates of the rate, the blood and liver cadmium are included in the model in our second-stage analysis. From Table 3, it is noted that the estimated R_k with low u_β_2 is lower than that with high u_β_2, which suggests the region of starting values for some unknown parameters, at the next step of estimation in the four-compartment model.

3.2.2. The second-stage analysis

In the four-compartment model, there are six unknown parameters C_{bl}, C_{lb}, C_{bk}, C_{kb}, C_{ku}, and R. The required optimization is very complicated in this case of the four-compartment model, since the number of responses and the number of unknown parameters are rather large. One of the most important things to ensure a successful nonlinear analysis is to obtain good starting values for the unknown parameters, before implementing the estimation procedure. There are two techniques used for determining the starting values of unknown parameters in this four-compartment model, viz. a graphical analysis and a grid search.

With the graphical analysis, the meaning of each parameter can be interpreted graphically to describe the behavior of the expectation functions. There are three references, the half-life times of liver and kidney cadmium and blood compartment with a rapid turnover from previous results, the portions of body burden of kidney and liver cadmium from previous results, and the estimates of excretion rate and cadmium intake from the two-compartment model discussed earlier, forming the basis of interpreting the unknown parameters $[C_{bl}, C_{lb}, C_{bk}, C_{kb}, C_{ku}, R]'$.

The steps involved in the graphical analysis are as follows.

1. Solve the expectation functions numerically after fixing the parameters within reasonable ranges based on the meaning of the parameters just mentioned above; Matlab software package was used for obtaining the numerical solutions, using Runge-Kutta-Fehlberg iteration method.

2. Draw the sketches of both fitted values derived from the numerical solution and the observed values of blood cadmium against time, liver cadmium against time, kidney cadmium against time, and urine cadmium against time, simultaneously, and visualize them.
3. Change the values of fixed parameters and repeat steps (1)–(3) until the fitted and observed values of blood cadmium, liver cadmium, kidney cadmium, and urine cadmium are reasonably close.

With the grid search, a combination of all resources for the graphical analysis and the results of the graphical analysis were used as references for the range of starting values of the parameters. In attempting to find proper starting values for the grid search, there are many steps taken. In the beginning, the increments are quite large for a rough search; then the increments become smaller and smaller for a finer search each time. The program written in C is intended not to give final results automatically, but to provide middle results step by step. During the search, it can be visualized on screen in order to give an idea of which part of the range is most likely to be selected. Moreover, the program can be paused at any time in the middle of searching if it is necessary, for deciding about adjusting the increments or even changing the range of parameters. The results of the grid search for the starting values of the parameters are shown in Table 4. Although this type of searching is very time consuming, it provides a general idea of where the location of the local minimum might be obtained within a reasonable range of the parameters. Furthermore, it most certainly accelerates the convergence of the solution of the nonlinear optimization problem.

Table 4. Results of the starting values for the parameters obtained from grid search for the four-compartment model.

	$\hat{C}_{bl}$	$\hat{C}_{lb}$	$\hat{C}_{bk}$	$\hat{C}_{kb}$	$\hat{C}_{ku}$	$\hat{R}$	O.F.V.
Subject 1	22	.002	7	.017	.007	414	30.5
Subject 2	7.8	.0033	13.1	.0045	.017	330	17.83
Subject 3	17	.043	18.4	.0092	.018	141	25
Subject 4	13	.0001	12	.0038	.008	512.5	19.96
Subject 5	0.2	.0019	14.2	.0001	.0105	121.1	36.6327
Subject 10	17	.007	1.8	.0004	.0116	530	154.81
Subject 11	9.6	.0059	19.2	.015	.027	1265	9.3449
Subject 13	8.4	.0013	11.2	.0035	.03	840.7	100.9133

Table 5 shows the results of the four-compartment model based on the nonlinear least-squares estimation, while Table 6 shows the results of the four-compartment model from the Bayesian approach based on simulated and observed data. The steps of generating the simulated data are similar to those based on the two-compartment model. Estimates of C_{bl}, C_{lb}, C_{bk}, and C_{kb} from both approaches provide some information about the blood distribution of cadmium to liver and kidney and the cadmium elimination rate from blood either to liver or to kidney, which the two-compartment model was unable to provide. By comparing the estimates of C_{ku} in the four-compartment model and of λ_k in the two-compartment model, it is noticed that the results from the two-compartment model and the four-compartment model remain close. However, the estimate of R_k in the two-compartment model differs greatly from the estimate of R in the four-compartment model. The convergence of the optimization procedure goes smoothly due to the fact that the grid search provides good starting values for the unknown parameters.

All subjects except 13 show that the fitted data from the four-compartment model agree reasonably well with the observed data. u_β_2 for Subject 13 is very high, which indicates that

Table 5. Results of the nonlinear least-squares estimation method for the four-compartment model.

Subject	$\hat{C}_{bl}$	$\hat{C}_{lb}$	$\hat{C}_{bk}$	$\hat{C}_{kb}$	$\hat{C}_{ku}$	$\hat{R}$	O.F.V.
1	21.98	.00001	6.993	.0164	.0067	414.0	27.7
2	8.68	.0039	21.32	.013	.0169	325.03	16.7581
3	14.30	.0568	15.70	.0037	.0177	141.00	24.20
4	13.39	.00001	12.24	.0038	.0079	510.09	19.86
5	6.74	.0705	10.82	.0001	.0107	62.10	12.0684
10	17.25	.0075	1.15	.0002	.0116	530.0	154.11
11	12.02	.008	17.98	.013	.0273	1259.9	9.22
13	9.15	.0021	11.55	.0045	.0304	837.21	100.82

Table 6. Results for the four-compartment model from the Bayesian approach based on simulated and observed data.

Subject	$\hat{C}_{bl}$	$\hat{C}_{lb}$	$\hat{C}_{bk}$	$\hat{C}_{kb}$	$\hat{C}_{ku}$	$\hat{R}$	O.F.V.
2	10.33	0.0054	19.67	0.0116	0.0171	294.82	3216.74
4	14.05	0.000004	11.89	0.0033	0.0078	523.5	5668.0
11	11.46	0.0085	18.54	0.0166	0.0253	1174.09	79.44

the kidney functions for this subject may be damaged and some irregular and unpredicted pattern could have taken place in this case. Figures 3–5 are examples showing good agreement of the observed data with the fitted four-compartment model for Subjects 2, 4, and 11.

3.3. Discussions and Suggestions

The large estimates of C_{bl} and C_{bk} confirm that there is no cadmium accumulation in blood with rapid turnover of cadmium concluded from previous results.

A combination of C_{kb} and C_{ku} reflects the half-life time in the kidney. Absence of feces cadmium variable might cause the estimate of the half-life time in the kidney to be small. It is probable that missing observations of feces cadmium leads to the overestimation of C_{ku}.

It is noticed from the results that estimates of C_{lb}, C_{bl}, C_{bk}, C_{ku}, and R might be affected by the absence of observations of tissue cadmium and feces cadmium. The range of estimates of C_{lb} indicates the half-life time in liver. The estimate of C_{lb} seems small, which might be due to no tissue cadmium variable being involved in the model. It might affect the estimate of C_{lb} since cadmium may be transported back from tissue to liver via blood, and missing observations of tissue cadmium might lead to both C_{bl} and C_{bk} being overestimated and C_{lb} being underestimated. Unfortunately, there is no entrance to measure the tissue cadmium yet. Therefore, when the half-life time of blood, liver, and kidney cadmium are all derived from each of these estimates, the influence of the absent tissue cadmium should be taken into account.

Since the tissue cadmium is about 31% of the total body burden, it should be included in the model when the number of measurements of the blood cadmium and others increase to a certain extent, in spite of unavailable observation on the tissue cadmium. The half-life time of tissue cadmium and the relationship with the other compartments can be estimated through blood cadmium and other observations without tissue cadmium, based on the Bayesian approach, in a similar way as shown in Table 7. Table 7 gives the results of the four-compartment model from the Bayesian approach derived just from blood cadmium observations. It should be noticed that

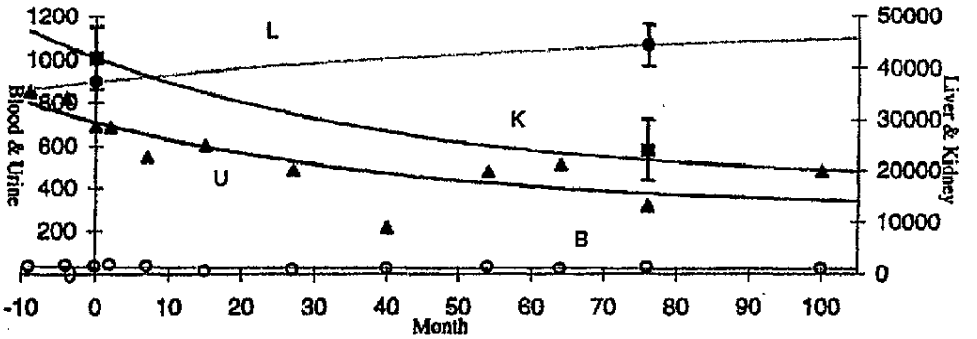

Figure 3. Plot of the observed data for Subject 2 with the fitted four-compartment model.

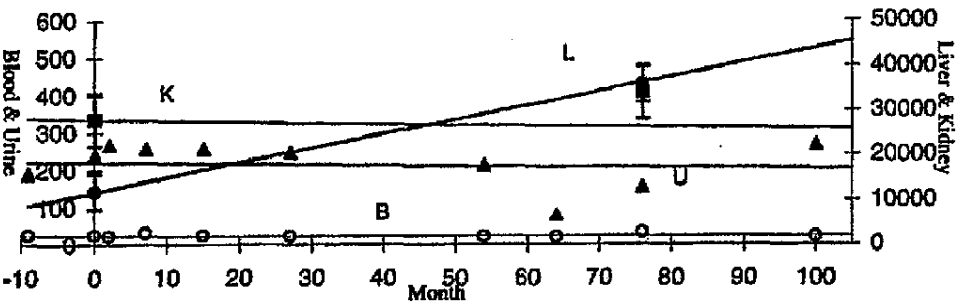

Figure 4. Plot of the observed data for Subject 4 with the fitted four-compartment model.

Table 7. Results for the four-compartment model from the Bayesian approach based only on blood cadmium observations.

Subject	$\hat{C}_{bl}$	$\hat{C}_{lb}$	$\hat{C}_{bk}$	$\hat{C}_{kb}$	$\hat{C}_{ku}$	$\hat{R}$	O.F.V.
1	22.0	.000003	7.00	.0177	.0071	414.0	4.24
2	7.8	.0039	13.10	.0070	.0531	330.0	6.29
3	14.3	.052	15.70	.0064	.0375	140.99	7.48
4	14.0	.000006	12.00	.0054	.0105	500	2.04
5	6.70	.0704	11.00	.0002	.000008	60	1.19
10	16.98	.006	1.79	0	.007	739	6.88
11	9.40	.013	16.79	.00001	.816	1265.0	1.77
13	8.40	.000004	11.2	.0051	.018	840.70	9.93

the results derived from blood cadmium alone are based on good starting values provided by the graphical analysis and the grid search based on all responses.

The estimate of R is quite reasonable. It is found from the results of the four-compartment model that subjects with high β_2-microglobulin have higher estimates of intake cadmium (R) than those with low β_2-microglobulin. The interpretation for this could be that the environment is not only the source for cadmium intake, but that the cadmium released from other tissues might also contribute to the intake of cadmium.

The differences of other estimates like C_{bl}, C_{lb}, C_{bk}, C_{kb}, and C_{ku} between the two subgroups are not obvious. It is not conclusive because of a small and insufficient data set and the uncertainty in observations; it perhaps implies that the distribution of cadmium in the body depending on the absorption and the metabolism of cadmium in the body varies from individual to individual,

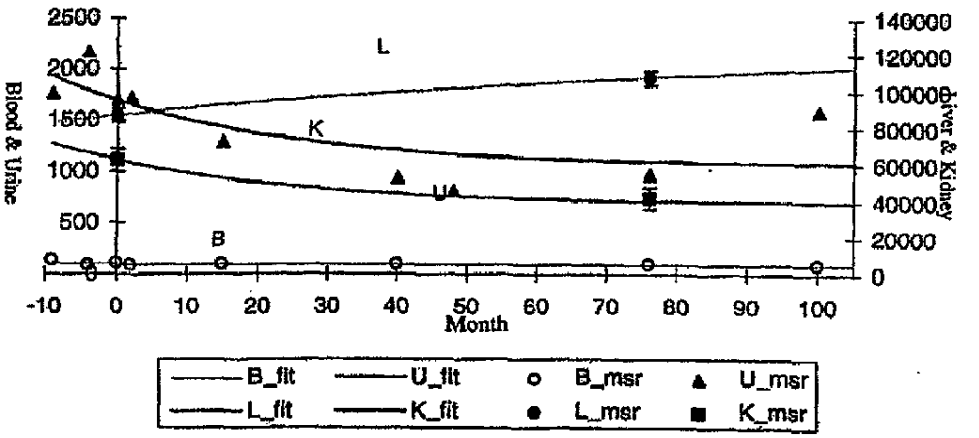

Figure 5. Plot of the observed data for Subject 11 with the fitted four-compartment model.

all that it is not statistically different between a worker with abnormal kidney condition and one with normal kidney condition.

In order to formulate further refined models and develop more powerful statistical analysis for the model, as the cadmium research still continues on actively, there are four suggestions for cadmium researchers. First, the number of subjects and the number of measurements should be increased in order to carry out appropriate statistical analysis effectively. Second, since it is more economical to collect blood and urine cadmium samples from workers exposed to cadmium than liver and kidney cadmium *in vivo* measurements, blood and urine samples should be gathered more often for preliminary analysis. A series of blood cadmium variables can be used for rough estimation of C_{bl}, C_{bk}, and R by a four-compartment model. Even though the other three parameters, C_{lb}, C_{kb}, and C_{ku}, can also be estimated through blood cadmium, the estimation might be poor for C_{lb} and C_{kb} (with very small values) and for C_{ku} (with no direct connection to blood cadmium), but they can still be a reference for researchers. In the meantime, a series of urine cadmium variables and a few kidney cadmium variables can be used to estimate R and C_{ku} by a two-compartment model. In addition, tissue cadmium variable can be incorporated into the model in order to estimate some rates relating to the tissue cadmium. Third, it may be possible in the future to investigate the cadmium metabolism in the human body in stratified subgroups with kidney condition when sufficient data are available. Moreover, nonlinear segmented models according to different exposure levels in different time periods can be formulated. Consequently, several statistical analyses may be needed based on different time intervals. The model should be formulated on the data with shorter time intervals at first, and then designed for the data with longer and longer time intervals. Finally, after the completion of *in vivo* measurements, entire analysis can be done using a four-compartment model through the Bayesian approach. It is also possible to analyze the cadmium metabolism by a five-compartment model including the tissue cadmium, provided sufficient *in vivo* measurements of liver and kidney cadmium are available. Since different responses in the same case would have been collected from the same body, correlation will exist between them. Therefore, we strongly recommend the use of the Bayesian determinant minimization method for the cadmium analysis whenever sets of such multiresponse data are available.

REFERENCES

1. R. Armstrong, D.R. Chettle, M.C. Scott, M. Blindt and H.J. Mason, Longitudinal studies of exposure to cadmium, *British Journal of Industrial Medicine* **49**, 556–559, (1992).
2. R. Armstrong, R.A. Braithwaite, D.R. Chettle, M.C. Scott and D.M. Franklin, A cross-sectional and longitudinal survey of workers exposed to cadmium, Report presented at the *Workshop on In Vivo Body Composition Studies*, Swansea, Wales, (1991).

3. T. Kjellström and G.F. Nordberg, A kinetic model of cadmium metabolism in the human being, *Environmental Research* **16**, 248–269, (1978).
4. J.A. Jacquez, *Compartmental Analysis in Biology and Medicine*, Elsevier, Amsterdam, (1972).
5. A.P. Mishina and I.V. Proskuryakov, *Higher Algebra*, Pergamon Press, London, (1965).
6. R.K. Miller and A.N. Michel, *Ordinary Differential Equations*, Academic Press, London, (1982).
7. G.E.P. Box and G.C. Tiao, *Bayesian Inference in Statistical Analysis*, John Wiley & Sons, New York, (1973).
8. D.M. Bates and D.G. Watts, *Nonlinear Regression and Its Applications*, John Wiley & Sons, New York, (1988).